

Julissa Mijares  
 Professor Kontothanassis  
 CDS DS 210  
 6 May 2023

## Final Project Write-Up

Answer the question: Degree distributions: What is the distribution of vertex degrees in your graph? What if you look at the number of neighbors at distance 2? Often people believe that such a distribution should be a power-law distribution (look up online what it means) for social networks and similar graphs. How well does the distribution you see fit this assumption (and how are you going to evaluate it)? Degree of a vertex is the number of other vertices it is connected to.

- Based on my degree\_distribution.png, my distribution does closely follow a power-law distribution.

### Dataset:

In this project, I will be analyzing the "twitter\_combined.txt" dataset, which represents a social network of Twitter users, where each line represents an edge between two nodes. The dataset consists of 81,306 nodes and 1,768,149 edges. The dataset can be downloaded from <https://snap.stanford.edu/data/ego-Twitter.html>. The goal is to analyze the distribution of vertex degrees in the graph and evaluate whether it fits the power-law distribution.

### Project Description:

The project is implemented in Rust programming language and divided into three modules to keep it organized and easy to read: main.rs, graph.rs, and plot.rs.

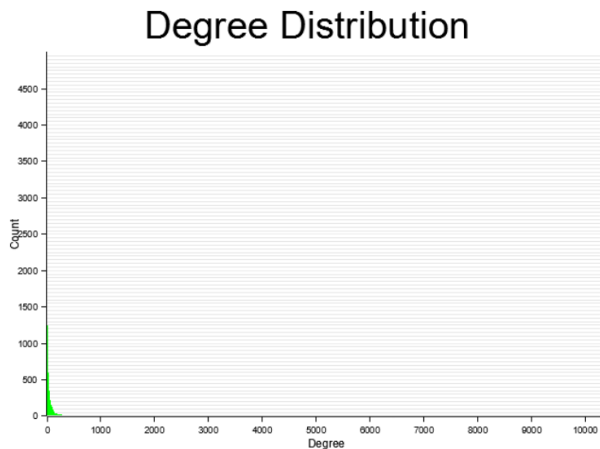
- In the graph.rs module, the Graph struct is defined, which represents the graph data structure. The Graph struct has methods to add an edge between two nodes, get the degree of a node, and get the set of nodes that are neighbors of a node at a distance of 2 in the graph.
- In the plot.rs module, the plot\_degree\_distribution function is defined, which takes a vector of degrees. The function first calculates the frequency of each degree in the vector, then generates a histogram of the degree distribution using gnuplot, and saves it to a file named "degree\_distribution.png".
- In the main.rs module, the graph data is read from the "twitter\_combined.txt" file, and a new Graph instance is created. The degree distribution of the graph is calculated using the plot\_degree\_distribution function from the plot.rs module. The plot\_degree\_distribution function takes a vector of degrees and generates a histogram of the degree distribution.

Algorithms: The project uses two algorithms to analyze the Twitter graph:

1. The degree algorithm: The degree algorithm is used to calculate the degree of a node in the graph. The degree of a node is the number of other nodes it is connected to.
2. The neighbor-at-distance-2 algorithm: The neighbor-at-distance-2 algorithm is used to find all nodes that are neighbors of a given node in the graph. The algorithm first finds the set of adjacent nodes to the given node and then finds the set of neighbors of those adjacent nodes that are not the given node itself.

### Output:

The output of running the project is a histogram of the degree distribution of the Twitter graph saved in a file named "degree\_distribution.png". Additionally, the program outputs the set of nodes that are neighbors of node 0 in the graph. The output will be printed in the console and saved in the output directory.



```
Finished dev [unoptimized + debuginfo] target(s) in 2.07s
Running target\debug\ds210_final.exe
Result has been saved to degree_distribution.png
Successfully plotted degree distribution
Neighbors at distance 2 of node 0: {}
```